7C

УДК 004.93

С.А. Субботин
Запорожский национальный технический университет, Украина
subbotin@zntu.edu.ua

Неитеративный синтез нейро-нечетких диагностических экспертных систем

Предложен метод синтеза диагностических экспертных систем на основе нейро-нечетких моделей, позволяющий использовать априорные экспертные знания, автоматически формировать разбиение признакового пространства и выделять нечеткие термы на основе эмпирических наблюдений, а в неитеративном режиме – определять значения весовых коэффициентов, функции постсинаптического потенциала и активационные функции нейро-нечеткой сети.

Введение

Уровень развития промышленности передовых стран мира наряду с объемом производства и ассортиментом выпускаемой продукции характеризуется также показателями ее качества, надежности и эксплуатационной безопасности. Поэтому для отечественной промышленности, интенсивно интегрирующейся в мировую экономику, особую актуальность приобретает управление качеством изделий, среди наиболее важных инструментов которого, прежде всего, следует выделить неразрушающую диагностику [1].

Одним из наиболее распространенных подходов при построении диагностических систем является применение методов распознавания образов и искусственных нейронных сетей [1], [2]. Недостатками таких сенсорных систем является невозможность или крайне высокая сложность ввода экспертных знаний в систему, высокая итеративность процесса построения диагностической модели, зависимость от пользователя при определении начальных параметров модели.

Другим традиционным подходом для автоматизации диагностики являются экспертные системы, основанные на знаниях [3], которые позволяют легко вводить и использовать знания, формализованные экспертами, однако не обладают возможностями обучаться в процессе построения и эксплуатации по наблюдениям.

Поэтому актуальной является разработка и исследование методов, обеспечивающих объединение экспертных знаний и эмпирических наблюдений для принятия решений.

Весьма перспективным классом средств, объединяющих достоинства сенсорных и знаниеориентированных систем являются нейро-нечеткие сети [4-7], обладающие способностями к обучению и позволяющие вводить в структуру сети априорные экспертные знания. Тем не менее, для большинства известных методов построения нейро-нечетких диагностических моделей характерны проблемы задания числа нечетких термов для диагностических признаков, а также итеративность оптимизационного подбора параметров нечетких термов и неопределенность выбора их начальных знаний.

Целью данной работы являлось создание метода построения нейро-нечетких диагностических экспертных систем, позволяющего устранить проблемы настройки параметров нечетких термов, а также автоматически формировать нечеткие термы по эмпирическим данным.

Постановка задачи. Пусть задана обучающая выборка <x, y>, где $x = \{x^s\}$, $y = \{y^s\}$, $x^s = \{x^s_j\}$, x^s_j – значение j-го признака s-го экземпляра выборки, y^s – значение выходного признака, сопоставленное s-му экземпляру обучающей выборки, $y^s \in \{k\}$, S = 1, 2, …, S; j = 1, 2, …, N; k =1, 2, …, K; S – количество экземпляров, N – количество признаков, K – число классов.

Для построения диагностической нечеткой экспертной системы необходимо сформировать разбиение признакового пространства на кластеры, для каждого кластера выделить нечеткие термы по признакам, определить схему нечеткого вывода номера класса из значений функций принадлежности к нечетким термам, представить нечеткую систему в нейробазисе.

Учет экспертных знаний

Экспертные знания удобно представить набором правил вида:

Если $\bigwedge_{j=1}^{N}\left(x_j^g \in [l_{j,q}, r_{j,q}]\right)$, то $y^g = k$ с коэффициентом доверия α_g,

где g – номер экспертного правила, g = 1, 2, …, G; $l_{j,q}$ и $r_{j,q}$ – левая и правая границы q-го интервала значений j-го признака соответственно; $\alpha_g \in [0,1]$.

После этого целесообразно проверить дублирование и противоречивость экспертных правил. Из нескольких одинаковых правил следует оставить только одно правило. Если одинаковые правила имеют разные коэффициенты доверия, то следует оставлять правило с наименьшим коэффициентом доверия. Из нескольких правил с одинаковыми антецедентами, но разными консеквентами, следует оставить то правило, которое имеет наибольший коэффициент доверия.

Для результирующего набора экспертных правил скорректировать значение G, определить количество интервалов значений по каждому признаку g_j, занести в $B_{g,j}$ номер интервала (терма) j-го признака, входящий в антецедент g-го правила (g = 1, 2, …,G; j = 1, 2, …, N), а также установить номер класса: $K_g = y^g$, g = 1, 2, …, G.

Формирование разбиения пространства признаков

Разбиение признакового пространства для выборки эмпирических наблюдений необходимо для определения нечетких термов признаков как проекций соответствующих блоков на координатные оси.

Формирование разбиения предлагается осуществлять путем выполнения последовательности шагов 1 – 12.

Шаг 1. Инициализация. Задать обучающую выборку <x, y>.

Шаг 2. По оси каждого признака j = 1, 2, …, N определить одномерные расстояния между экземплярами:

$$R_j(s,p) = \left|x_j^s - x_j^p\right|.$$

Среди полученных расстояний найти минимальное расстояние, большее нуля:

$$R_j = \min_{\substack{s=1,2,\ldots,S;\\ p=s+1,\ldots,S}} \{R_j(s,p) \mid R_j(s,p) > 0\},\ j = 1, 2, \ldots, N.$$

Шаг 3. Для каждого признака определить количество интервалов разбиения диапазона его значений:

$$n_j = \frac{\max(x_j) - \min(x_j)}{R_j},\ j = 1, 2, \ldots, N,$$

а также определить длину интервала наблюдаемых значений каждого признака:

$$r_j = \max(x_j) - \min(x_j),\ j = 1, 2, \ldots, N.$$

Шаг 4. Разбить ось j-го признака на n_j интервалов. Определить координаты левых и правых границ для каждого p-го интервала j-го признака по формулам:

$$l_{j,p} = \min(x_j) + (p-1)\frac{r_j}{n_j},\ \ r_{j,p} = \min(x_j) + p\frac{r_j}{n_j},$$

$$j = 1, 2, \ldots, N;\ p = 1, 2, \ldots, n_j.$$

Шаг 5. Сформировать блоки-кластеры и задать номера их классов путем выполнения шагов 5.1 – 5.8.

Шаг 5.1. Сформировать прямоугольные блоки $\{B_q\}$, $q = G+1, G+2, \ldots, G+Q$, $Q = \prod_{j=1}^{N} n_j$, в N-мерном пространстве признаков на пересечении соответствующих интервалов значений признаков. Занести в $B_{q,j}$ номер интервала j-го признака, который соответствует q-му блоку.

Шаг 5.2. Определить номера классов для прямоугольных блоков в N-мерном пространстве признаков:

$$K_q = \begin{cases} \{k \mid y^s = y^t = k, l_{j,B_{q,j}} \le x_j^s \le r_{j,B_{q,j}}, l_{j,B_{q,j}} \le x_j^t \le r_{j,B_{q,j}}, \\ s = 1,\ 2, \ldots, S, t = s+1, \ldots, S;\ j = 1,\ 2,\ \ldots, N; k = 1,\ 2,\ \ldots, K\}; \\ 0, \neg\exists s = 1,\ 2, \ldots, S : l_{j,B_{q,j}} \le x_j^s \le r_{j,B_{q,j}}, j = 1,\ 2,\ \ldots, N; \\ -1, \text{в противном случае.} \end{cases}$$

Установить коэффициент уверенности классификации для блоков:

$$\alpha_q = \begin{cases} 1, \text{если } K_q > 0; \\ 0, \text{если } K_q \le 0. \end{cases}$$

Шаг 5.3. Для тех блоков, у которых $K_q = -1$, $q = G+1, G+2, \ldots, G+Q$, установить:

$$K_q = \arg \max_{k=1,2,\ldots,K} S_q^k,$$

где S^k_q – количество экземпляров k-го класса, попавших в q-й блок-кластер.

Шаг 5.4. Для тех блоков, у которых номер класса $K_q = 0$, $q = G+1, G+2, \ldots, G+Q$, определить расчетный номер класса, для чего предлагается использовать модифицированный нерекуррентный метод потенциальных функций [1].

Шаг 5.5. Вычислить расстояние между q-м и p-м блоками, $q = G+1, G+2, \ldots, G+Q$, $p = q+1, \ldots, G+Q$, как:

$$R(B_q, B_p) = \sqrt{\sum_{j=1}^{N} r_j^{-1}\left(C_{q,j} - C_{p,j}\right)^2} \ \text{ или } \ R(B_q, B_p) = \sqrt{\sum_{j=1}^{N} r_j^{-1}\left|C_{q,j} - C_{p,j}\right|},$$

где $C_{q,j} = \dfrac{l_{j,B_{q,j}} + r_{j,B_{q,j}}}{2}$. Заметим, что $R(B_q, B_p) = R(B_p, B_q)$.

Шаг 5.6. Определить потенциал, наводимый совокупностью блоков, принадлежащих к k-му классу, на p-й блок с неизвестной классификацией:

$$\Phi_p^k = \frac{1}{L_k}\sum_{q=1}^{Q}\left\{S_q e^{-R^2(B_q,B_p)} \middle| K_q = k, K_p = 0, q \ne p\right\},\ q = G+1, G+2, \ldots, G+Q,\ p = q+1, \ldots, G+Q,$$

где L_k – количество блоков, принадлежащих к k-му классу, S_q – количество экземпляров обучающей выборки, попавших в q-й блок.

Шаг 5.7. Установить номер класса для p-го блока с неизвестной классификацией ($K_p = 0$) по формуле:

$$K_p = \arg \max_{k=1,2,\ldots,K} \Phi_p^k, p = G+1, G+2, \ldots, G+Q.$$

Шаг 5.8. Модифицировать значения коэффициентов уверенности для блоков:

$$\alpha_q = \{\Phi_q^{K_q} \mid \alpha_q = 0\}, \; q = G+1, G+2, \ldots, G+Q.$$

Шаг 6. Выполнить объединение смежных блоков-кластеров.

Выполнить объединение смежных блоков, принадлежащих к одному и тому же классу: для $\forall q, p = G+1, G+2, \ldots, G+Q;\; q \neq p$: если $K_q > 0$, $K_q = K_p$ и $\exists j : \left|B_{q,j} - B_{p,j}\right| = 1, \forall i \neq j : B_{q,i} = B_{p,i}, i = 1,2,\ldots,N, j = 1,2,\ldots,N;$ тогда объединить блоки q и p по j-му признаку:

– установить:

$$\alpha_q = \frac{\alpha_q \prod_{j=1}^{N}\left(r_{j,B_{q,j}} - l_{j,B_{q,j}}\right) + \alpha_p \prod_{j=1}^{N}\left(r_{j,B_{p,j}} - l_{j,B_{p,j}}\right)}{\prod_{j=1}^{N}\left(r_{j,B_{q,j}} - l_{j,B_{q,j}}\right) + \prod_{j=1}^{N}\left(r_{j,B_{p,j}} - l_{j,B_{p,j}}\right)},$$

$$n_j = n_j + 1,$$

$$l_{j,n_j} = l_{j,\min(B_{q,j},B_{p,j})},\; r_{j,n_j} = r_{j,\max(B_{q,j},B_{p,j})},\; B_{q,j} = n_j;$$

– удалить p-й блок: $K_p = 0$, $\alpha_p = 0$, $B_{p,i} = 0$, $i = 1,2, \ldots, N$.

Шаг 7. Из обучающего множества выделить подмножество экземпляров, относящихся к блокам-кластерам, номера классов которых не совпадают с номерами классов экземпляров. Применить для полученного разбиения и выделенного подмножества процедуру уточнения разбиения и дообучения модели.

Шаг 8. Останов.

Уточнение разбиения и дообучение модели

Если имеется разбиение признакового пространства, которое нужно уточнить (дообучить) на основе новых наблюдений $<x^*, y^*>$, $x^*=\{x^{s^*}\}$, $x^{s^*}=\{x^{s^*}{}_j\}$, $y^*=\{y^{s^*}\}$, $j = 1, 2, \ldots, N$; $s^* = 1, 2, \ldots, S^*$; то необходимо из набора новых наблюдений исключить те наблюдения, которые попадают в блоки имеющегося разбиения и соответствуют им по номеру класса, скорректировав соответствующим образом S^*. Для тех наблюдений, которые не совпадают с классами блоков, целесообразно сформировать отдельные точечные кластеры. Для каждого нового наблюдения сформировать интервалы по признакам и занести в $B_{G+z+s^*,j}$ номера интервалов для каждого j-го признака, соответствующие новому кластеру, а также определить:

$$l_{j,B_{G+z+s^*,j}} = x_j^{s^*} - \gamma R_j,\; r_{j,B_{G+z+s^*,j}} = x_j^{s^*} + \gamma R_j,\; K_{G+z+s^*} = y^{s^*},\; \alpha_{G+z+s^*} = 1,$$

$$s^* = 1, 2, \ldots, S^*; j = 1, 2, \ldots, N;$$

где γ – некоторая константа, $\gamma \in (0,1)$.

Синтез нейро-нечеткой экспертной системы

На основе полученных разбиений может быть синтезирована диагностическая экспертная система в виде нейро-нечеткой сети, схема которой изображена на рис. 1.

Важной особенностью сети является то, что она одновременно объединяет в себе три разбиения признакового пространства: на кластеры, сформированные на основе экспертных знаний (на рис. 1 выделены точками), на кластеры, сформированные на основе прямоугольного разбиения с использованием модификации нерекуррентного метода потенциальных функций по обучающей выборке (на рис. 1 выделены пунктиром), а также на кластеры, соответствующие точечным наблюдениям, номера классов которых не совпадают с номерами классов, к которым относятся блоки прямоугольного разбиения (на рис. 1 выделены штрих-пунктиром).

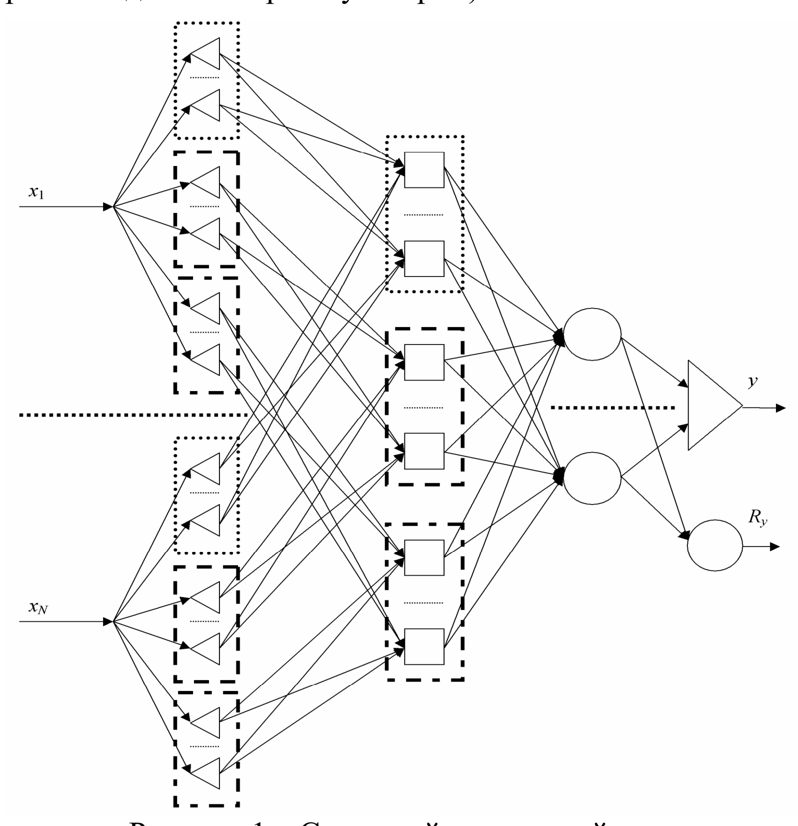


Рисунок 1 – Схема нейро-нечеткой сети

На входы сети подаются значения признаков распознаваемого экземпляра. Нейроны первого слоя вычисляют принадлежности распознаваемого экземпляра к термам признаков $\mu_{j,p}(x^s)$ (фаззификация). При этом по каждому признаку сначала определяются принадлежности к термам кластеров, сформированных на основе экспертных знаний, затем к термам кластеров, сформированных на основе прямоугольного разбиения, после чего к термам точечных наблюдений. Нейроны второго слоя определяют принадлежности к блокам-кластерам $\mu_q(x^s)$, которые также как и термы сгруппированы по разбиениям. Нейроны третьего слоя объединяют принадлежности к блокам в принадлежности к классам $\mu_k(x^s)$. После чего первый нейрон четвертого слоя осуществляет дефаззификацию результата, а второй нейрон четвертого слоя определяет достоверность классификации R_y.

Для расчета принадлежностей распознаваемого экземпляра x^s к нечетким термам признаков целесообразно использовать трапециевидные функции принадлежности:

$$\mu_{j,p}(x^s)=\begin{cases}0, x^s<l_{j,p};\\ \dfrac{x^s-l_{j,p}}{l_{j,p}(2-\delta)-l_{j,p}}, l_{j,p}\le x^s<l_{j,p}(2-\delta);\\ 1, l_{j,p}(2-\delta)\le x^s\le \delta r_{j,p};\\ \dfrac{r_{j,p}-x^s}{r_{j,p}-\delta r_{j,p}}, \delta r_{j,p}<x^s\le r_{j,p};\\ 0, r_{j,p}<x^s,\end{cases}$$

где δ – некоторая константа, $0<\delta\le 1$, либо треугольные функции принадлежности:

$$\mu_{j,p}(x^s)=\begin{cases}0, x^s<l_{j,p};\\ \dfrac{x^s-l_{j,p}}{0,5(r_{j,p}-l_{j,p})}, l_{j,p}\le x^s<0,5(r_{j,p}+l_{j,p});\\ \dfrac{r_{j,p}-x^s}{0,5(r_{j,p}-l_{j,p})}, 0,5(r_{j,p}+l_{j,p})<x^s\le r_{j,p};\\ 0, r_{j,p}<x^s.\end{cases}$$

Весовые функции нейронов сети будут задаваться формулами:

$$\varphi_j^{(2,i)}\left(w_j^{(2,i)},x_j^{(2,i)}\right)=\max\{w_j^{(2,i)},x_j^{(2,i)}\}, i=1, 2, \ldots, Q; j=1, 2, \ldots, V+z+NS^*;$$
$$\varphi_j^{(3,i)}\left(w_j^{(3,i)},x_j^{(3,i)}\right)=\min\{w_j^{(3,i)},x_j^{(3,i)}\}, i=1, 2, \ldots, K; j=1, 2, \ldots, G+Q+S^*;$$
$$\varphi_j^{(4,i)}\left(w_j^{(4,i)},x_j^{(4,i)}\right)=\min\{w_j^{(4,i)},x_j^{(4,i)}\}, i=1, 2; j=1, 2, \ldots, K;$$

где $V=\sum_{j=1}^{N}g_j$, $z=\sum_{j=1}^{N}n_j$, $x_j^{(\eta,i)}$ – значение сигнала на j-м входе i-го нейрона η-го слоя сети, $w_j^{(\eta,i)}$ – вес j-го входа i-го нейрона η-го слоя сети; $\varphi_j^{(\eta,i)}$ – весовая функция j-го входа i-го нейрона η-го слоя сети; g_j – количество интервалов значений (термов) j-го признака для набора экспертных правил.

Функции активации нейронов будут определяться по формулам:

$$\psi^{(2,i)}\left(\varphi_j^{(2,i)}(w_j^{(2,i)},x_j^{(2,i)})\right)=\min\{\varphi_j^{(2,i)}(w_j^{(2,i)},x_j^{(2,i)})\}, i=1, 2, \ldots, Q; j=1, 2, \ldots, V+z+NS^*;$$
$$\psi^{(3,i)}\left(\varphi_j^{(3,i)}(w_j^{(3,i)},x_j^{(3,i)})\right)=\max\{\varphi_j^{(3,i)}(w_j^{(3,i)},x_j^{(3,i)})\}, i=1, 2, \ldots, K; j=1, 2, \ldots, G+Q+S^*;$$
$$\psi^{(4,1)}\left(\varphi_j^{(4,1)}(w_j^{(4,1)},x_j^{(4,1)})\right)=\arg\max_j\{\varphi_j^{(3,i)}(w_j^{(3,i)},x_j^{(3,i)})\};$$
$$\psi^{(4,2)}\left(\varphi_j^{(4,2)}(w_j^{(4,2)},x_j^{(4,2)})\right)=\max_j\{\varphi_j^{(4,2)}(w_j^{(4,2)},x_j^{(4,2)})\},$$

где $\psi^{(\eta,i)}$ – функция активации i-го нейрона η-го слоя сети.

Весовые коэффициенты нейронов сети предлагается задавать по формуле:

$$w_j^{(\eta,i)}=\begin{cases}1, \eta=2, B_{i,p}\ne j, i=1,2,\ldots,G+Q+S^*, p=1,2,\ldots,N, j=1,2,\ldots,V+z+NS^*;\\ 0, \eta=2, B_{i,p}=j, i=1,2,\ldots,G+Q+S^*, p=1,2,\ldots,N, j=1,2,\ldots,V+z+NS^*;\\ \alpha_j, \eta=3, i=K_j, i=1,2,\ldots,K, j=1,2,\ldots,G+Q+S^*;\\ 0, \eta=3, i\ne K_j, i=1,2,\ldots,K, j=1,2,\ldots,G+Q+S^*;\\ 1, \eta=4, i=1,2, j=1,2,\ldots,K.\end{cases}$$

Как видно из приведенных формул, синтез и настройка параметров нейро-нечеткой сети осуществляются в неитеративном режиме, что позволяет избежать необходимости расчета производных целевой функции по весам сети, а также итеративного характера коррекции весов, присущего традиционно применяемым градиентным методам обучения на основе техники обратного распространения ошибки.

Заключение

В работе решена актуальная задача разработки диагностических экспертных систем для решения задач неразрушающего контроля качества, интегрирующих экспертные знания и эмпирические наблюдения.

Впервые предложен метод синтеза диагностических экспертных систем на основе нейро-нечетких моделей, позволяющий автоматически формировать разбиение признакового пространства, выделять нечеткие термы, в неитеративном режиме определить значения весовых коэффициентов, функции постсинаптического потенциала и активационные функции нейро-нечеткой сети, исключить необходимость вычисления производных целевой функции по весам сети и избежать проблемы выбора начальной точки поиска, что, в свою очередь, позволяет ускорить процесс получения диагностической нейро-нечеткой модели.

Разработанный метод может быть рекомендован для использования на практике в задачах неразрушающей диагностики и управления качеством.

Литература

1. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей : Монография / [В.И. Дубровин, С.А. Субботин, А.В. Богуслаев, В.К. Яценко]. – Запорожье : ОАО «Мотор – Сич», 2003. – 279 с.
2. Дубровін В.І. Методи оптимізації та їх застосування в задачах навчання нейронних мереж : навчальний посібник / В.І. Дубровін, С.О. Субботін. – Запоріжжя : ЗНТУ, 2003. – 136 с.
3. Субботін С.О. Подання й обробка знань у системах штучного інтелекту та підтримки прийняття рішень : навчальний посібник / С.О. Субботін. – Запоріжжя : ЗНТУ, 2008. – 341 с.
4. Abraham A. Neuro-Fuzzy Systems: State-of-the-Art Modeling Techniques / A. Abraham // Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence / Eds. : J. Mira and A. Prieto. – Granada : Springer-Verlag, 2001. – P. 269-276.
5. Борисов В.В. Нечеткие модели и сети / Борисов В.В., Круглов В.В., Федулов А.С. – М. : Горячая линия-Телеком, 2007. – 284 с.
6. Митюшкин Ю.И. Soft Computing: идентификация закономерностей нечеткими базами знаний / Митюшкин Ю.И., Мокин Б.И., Ротштейн А.П. – Винница : Универсум-Винница, 2002. – 145 с.
7. Ротштейн А.П. Интеллектуальные технологии идентификации: нечеткая логика, генетические алгоритмы, нейронные сети / А.П. Ротштейн. – Винница : УНИВЕРСУМ-Винница, 1999. – 320 с.

С.О. Субботін
Неітеративний синтез нейро-нечітких діагностичних експертних систем
Запропоновано метод синтезу діагностичних експертних систем на основі нейро-нечітких моделей, що дозволяє використовувати апріорні експертні знання, автоматично формувати розбиття простору ознак і виділяти нечіткі терми на основі емпіричних спостережень, а в неітеративному режимі – визначати значення вагових коефіцієнтів, функції постсинаптичного потенціалу й активаційні функції нейро-нечіткої мережі.

S.A. Subbotin
The Neuro-fuzzy Diagnostic Expert System Uniterated Synthesis
The method of diagnostic expert system synthesis based on neuro-fuzzy model is offered. It allows to use an a priori expert knowledge, to form automatically a partition of feature space and to select fuzzy terms on the basis of empirical observations, and also to determine weight values, weight and activation functions of neuro-fuzzy network in a uniterated mode.

Статья поступила в редакцию 03.04.2009.